

Spatial variability of soil lipids reflects vegetation cover in a French peatland

Renata Zocatelli^{a,b,c*}, Jérémy Jacob^{a,b,c}, Sébastien Gogo^{a,b,c}, Claude Le Milbeau^{a,b,c}, Julien Rousseau^d,

Fatima Laggoun-Défarge^{a,b,c}

^a Université d'Orléans, ISTO, UMR, 7327, 45071 Orléans, France

^b CNRS/INSU, ISTO, UMR 7327, 45071 Orléans, France

^c BRGM, ISTO, UMR 7327, BP 36009, 45060 Orléans, France

^d Sologne Nature Environnement, Parc de Beauvais – BP136, Romorantin-Lanthenay, France

ABSTRACT

The purpose of this study is to show how current vegetation in a peatland is imprinted in the lipid fraction of the underlying soil. La Gnette is a fen peatland located in Central France dominated by *Sphagnum* spp. and ericaceous shrubs, colonized by sedges (*Molinia caerulea*) and trees since the 1970s (*Betula pendula* and *Pinus sylvestris*). Lipid were exhaustively identified and quantified in the most abundant plants and in samples collected at the base of *M. caerulea* tussocks. Among relevant biomarkers, tricyclic diterpenes and methoxy-serratenes stand out as specific of *P. sylvestris*, betulin derivatives as specific of *B. pendula*, and some pentacyclic triterpene ketones and acetates as biomarkers of Ericaceae. Multivariate analysis applied to biomarker concentrations in soil samples from several sites permitted distinguishing three different vegetation types: vegetation cover: (i) closed vegetation dominated by *P. sylvestris*; (ii) closed vegetation dominated *B. pendula* and (iii) open vegetation or semi-open vegetation (early colonization by trees) constituted mainly by Ericaceae, *Sphagnum* and graminoids. The comparison of tree-specific biomarker concentrations with estimates of tree biomass allowed establishing quantitative relationships that were valid over at least 80 m² around sampling site. Although preliminary, these results are promising in attempting relating biomarker concentrations in geological archives to paleo-biomasses.

Keywords: spatial variability, surface peat, molecular biomarkers, vascular plant invasion

* Corresponding author. E-mail address: renata.zocatelli@cnrs-orleans.fr (R. Zocatelli).

1. Introduction

In recent history, many peatlands in Europe have experienced vascular plant invasion (Hansen, 1976; Berendse et al., 1994; Chambers et al., 1999; Jacquemyn et al., 2005). The increased abundance of trees such as *Betula spp* and *Pinus spp*, and gramineae such *Molinia caerulea* could be driven by natural dynamics and/or induced by direct and indirect human activities (fire, drainage, fertilization and climate change; Gimingham, 1972; Hunt et al., 1991, 1993; Boeye et al., 1999; Tomassen et al., 2003, 2004; Aerts et al., 1992; Woodin and Farmer, 1993; Hogg et al., 1995). Increased abundance of vascular plants has the potential to profoundly influence peatland biodiversity, hydrology, biogeochemistry, and landscape evolution (Bobbink and Lamers, 2002; Binet et al., 2013 and Gogo et al., 2011). However, although vegetation change in peatlands is widely recognized, there is a promising, if little explored, avenue for understanding their impacts on ecosystem processes (Limpens et al., 2008).

Several studies showed a link between biomarkers imprints in soil and plant diversity in the surrounding (Marseille et al., 1999; Trendel et al., 2010; González-Pérez et al., 2011; Zocatelli et al., 2012; Lavrieux et al., 2012; Ronkainen et al., 2013). In addition, the analysis of soil lipids revealed that biomarker imprints of soils can distinguish between pasture/grassland and forest (Zocatelli et al., 2012) and that the soil chemical legacy of former land uses can be preserved for decades (Lavrieux et al., 2012). As remarked by Eglinton and Eglinton (2008), little attention has been paid to the parameters that control the spatial variability of biomarkers, i.e. how do they reflect spatial heterogeneities and what is the current knowledge on their production, transport and sedimentation? In this vast new field of investigation, our study aimed at testing whether relationships between plant diversity and soil chemodiversity are still valid, and are quantitative, at a small scale ($1/10^{\text{th}}$ of a m^2) by

attempting to connect (i) the spatial variability of soil lipid imprints in peat with the surrounding vegetation, and (ii) the concentration of tree biomarkers with tree biomass. La Guette peatland, located in the southeast part of the French Region Centre is particularly well adapted to the study due to the low plant diversity, which is expected to provide precise relationships between soil lipids and their plant source. Our working hypothesis was that the lipid composition of a surface peat would be directly affected by the lipid composition of plants or groups of plants that developed in the immediate vicinity. To test this hypothesis, we analysed the biomarker imprint of soils sampled at several stations under distinct vegetation cover and tree density and compared it with the lipid imprints in possible plant sources.

2. Material and methods

2.1. Setting

La Guette is a peatland situated in Neuville-sur-Barangeon (Sologne, Cher) in the French Centre Region [154 m above sea level (a.s.l.); N 47°19', E 2°16'; Fig. 1)]. La Guette peatland covers 25 ha and is a transitional fen (pH ca. 4-4.5). The original peatland vegetation is dominated by *Sphagnum* spp. (majoritarian *S. rubellum*) and ericaceous shrubs - *Erica tetralix* and *Calluna vulgaris*. Aerial photographs (1944, 1952, 1966, 1973, 1976, 1983, 1999 and 2005) reveal (i) progressive afforestation of the site by *Pinus sylvestris* and *Betula pendula* up to a fire event in the mid-70s. This event re-opened the site and reset the vegetation. Since that time, a progressive afforestation is noted with, again, the development of *P. sylvestris* and *B. pendula*. Since the fire event, *Molinia caerulea* (Poaceae) spread in almost all parts of the site and developed tussocks. Tree species such as *P. sylvestris* and *B. pendula* are more sporadically distributed, but are closing the ecosystem (Fig. 1).

2.2. Sampling

Nine stations were selected on a NE-SW transect in order to cover the entire length of the peatland, to be representative of the main vegetation types, and to reflect gradients of afforestation (Fig. 1b). At each station, one soil sample (1.5 cm width x 10 cm length) was taken in horizontal direction at the base of 15cm high *M. caerulea* tussocks, above the peatland topsoil (Fig. 1c). Considering that tussocks only developed after the fire event, our samples are supposed to record vegetation changes that occurred during the last 40 yrs that is essentially the result of recent afforestation by *P. sylvestris* and *B. pendula*.

Because *M. caerulea* tussocks are essentially constituted by *M. caerulea* material, the background lipid signal is expected to be similar in all samples whereas variations in biomarker imprints are expected to be only related to the specific vegetation surrounding sampling points. Aerial parts of the six most abundant plants, *S. rubellum*, *E. tetralix*, *C. vulgaris*, *M. caerulea*, *P. sylvestris* (needles and bark) and *B. pendula* (leaves and bark), were collected as reference samples.

2.3. Tree density and dominance of *P. sylvestris* vs. *B. pendula*

Around each station, trees were counted in concentric circles of increasing radius (1, 2, 3, 4 and 5 m, giving areas of 3.14; 12.6; 28.3; 50.3 and 78.5 m², respectively). A biomass index (BI) is calculated for each area as the sum of trees height divided by the area. This was individually done for *P. sylvestris* and *B. pendula*, and for both tree species (formulae a, b and c, respectively). To give account of the relative abundance of these species, the biomass proportion of each to the total tree biomass was also calculated (formulae d and e, respectively).

$$BI_P = \Sigma \text{Pinus height in surface S} / \text{surface S} \quad (a)$$

$$BI_B = \Sigma \text{Betula height in surface S} / \text{surface S} \quad (b)$$

$$BI_T = \Sigma \text{all trees height in surface S} / \text{surface S} \quad (c)$$

$$P\% = \text{percentage of } P. \textit{sylvestris} \text{ biomass } (BI_P / BI_T \times 100) \quad (d)$$

$$B\% = \text{percentage of } B. \textit{pendula} \text{ biomass } (BI_B / BI_T \times 100) \quad (e)$$

2.4. Lipid analysis

About 5 g of fresh soil or lyophilized plant samples were ultrasonically extracted (3x, 10 min) with CH₂Cl₂ (15 ml). After combination of the extracts, the total extract was dried under N₂ and fractionated into neutral and acidic compounds using solid phase extraction with aminopropyl-bonded silica according to Jacob et al. (2005). Neutral and acidic fractions were dried under N₂. The neutral fraction was separated into: aliphatics and aromatics; ethers, esters, ketones and acetates; and alcohols using flash chromatography with a Pasteur pipette filled with silica (activated 24 h at 120 °C, then deactivated with H₂O, 5% wt.), using a sequence of solvents of increasing polarity. The acid fraction was methylated by adding a mixture of anhydrous MeOH and acetyl chloride kept at room temperature for 1 h. Alcohol and acid fractions were silylated by reacting with N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) in pyridine at 60 °C for 1 h. An internal standard, 5 α -cholestane, was added prior to analysis.

Lipids were assigned and quantified using gas chromatography-mass spectrometry (GC-MS) with a Trace GC Ultra gas chromatograph coupled to a TSQ Quantum XLS mass spectrometer equipped with an AS 3000 autosampler; both from Thermo-Scientific, Bremen, Germany). The GC instrument was fitted with a TG-5 MS column (60 m, 0.25 mm i.d., 0.25 μ m film thickness; Thermo, Bellefonte, PA, USA). The temperature program was: 40 °C (1 min) to 120 °C at 30 °C/min, then to 300 °C (held 70 min) at 3 °C/min. The sample was

dissolved in toluene and 2 μ l were injected in splitless mode at 280 °C. The carrier gas was He at 1.0 ml/min. The mass spectrometer was operated in the electron ionization (EI) mode at 70 eV and scanned from m/z 50 to 600. Compound assignment was based on comparison of mass spectra and retention times of authentic standards and on comparison with published data. Lipid concentration was estimated by measuring peak areas on ion specific chromatograms, converting to areas on total ion current (TIC) by applying a correction factor, then expressing values as a ratio of the area of the standard peak on the TIC, and finally normalized to the weight of sample extracted.

2.5. Statistical analysis

Model II regression (Legendre and Legendre, 1998) was used to examine the relationships between lipid concentration and tree density at each station. Pearson regression was used to examine the relationships between the regression coefficient of the previous analysis and the counting area. Statistical analysis was performed with Statistica (Statsoft, 2008) to test the significance of the regressions (level of significance set at $p < 0.05$). Molecular contents were compared through standardized principal component analysis (PCA).

3. Results

3.1. Cover vegetation parameters

Stations StI and StII are the most open ($BI_T = 0$; Table 1) as they have no trees within 5 m (78 m^2). Stations StIII, StV and StVIII are the most closed vegetation stations ($0.9 < BI_T < 1.53$; Table 1). StIII and StV are markedly dominated by *P. sylvestris* ($P_{\%} > 72\%$), whereas this species is very slightly dominating in StVIII ($P_{\%} = 51\%$). Stations StIV, StVI, StVII and

StIX are open to semi-open areas ($BI_T < 0.5$). Stations StIII, StV and StVIII are dominated by *P. sylvestris* ($P_{\%} > 51\%$), whereas *B. pendula* is dominant in stations StIV, StVI, StVII and StIX ($B_{\%} > 63\%$; Table 1).

3.2. Molecular imprints of soil samples

The distribution of main lipids identified and quantified in soil samples between stations (in terms of presence/absence), the ratios between homologous of *n*-alkyl lipids and lipid contents are shown in Tables 2, 3 and 4. Concentrations of each individual compound can be found in Supplementary data. The total amount of lipids ranged from ca. 54 to ca. 485.2 $\mu\text{g/g}$ soil.

n-Alkanes concentration ranged from 5.0 to 16.2 $\mu\text{g/g}$ soil, distributed from *n*-C₁₅ to *n*-C₃₅ with maxima varying between *n*-C₂₅, *n*-C₂₇, *n*-C₂₉ and *n*-C₃₃. All samples have a modal distribution, some maximizing at *n*-C₂₅, *n*-C₂₇ and *n*-C₂₉ (StII, StIII, StV, StVI and StVII), whereas others maximizing at *n*-C₃₁, *n*-C₃₃ (StI, StIV, StVIII and StIX, Table 3). The Carbon Preference Index (CPI) values for *n*-C₂₅ ~ *n*-C₃₃ *n*-alkanes ranged from 1.5 to 6.5. Much less variation was seen in the Average Chain Length (ACL, 27.6 – 30.5). The P_{aq} and *n*-C₂₅/*n*-C₂₉ ratio ranged from 0.25 to 0.57 and from 0.5 to 1.5, respectively. The $(n\text{-C}_{31} + n\text{-C}_{33})/(n\text{-C}_{27} + n\text{-C}_{29})$ ratio ranged from 0.4 to 3.3 (Table 3). *n*-Alkanols ranged from 0.2 to 38.0 $\mu\text{g/g}$ soil, distributed from *n*-C₁₉OH to *n*-C₃₁OH with maxima at *n*-C₂₈OH, with the exception of StIX that maximizes at *n*-C₂₆OH. *n*-Alkanols showed a strong even-over-odd predominance with a CPI generally ranging from 23.1 to 49.4, although StIX showed value of 4.7. ACL ranged from 23.4 to 27.8. The concentration of *n*-alkanoic acids ranged from 11.6 to 91.9 $\mu\text{g/g}$ soil distributed from *n*-C_{14:0} to *n*-C_{32:0} with maxima at *n*-C_{14:0} and *n*-C_{22:0}. Again,

there was a strong even-over-odd predominance with CPI and ACL values ranging from 2 to 7.8 and from 23.7 to 25.6, respectively (Table 3).

Selected compounds are depicted on the chromatograms of the ketone/acetate and alcohol fractions for several soil samples characterized by distinct vegetation cover (Fig. 2). Five steroids occurred in soil samples (st1 to st5). In all samples, β -sitosterol (st4) was the major compound, followed by campesterol (st2). Twelve pentacyclic triterpenones (k1 to k12), eight pentacyclic triterpenyl acetates (ta1 to ta7), eight pentacyclic triterpenols (tp1-8) and two pentacyclic triterpene acids (oleanoic and ursolic acids; Oa and Ua, respectively) were detected in soils samples (Table 2; see Supplementary Data for concentrations). Biomarkers [lupa-2,20(29)-dien-28-ol (db2), betulinic acid (db3), betulin (db4), tricyclic diterpenes (td1-6) and methoxy-serratenes (ms1 to ms5)] also detected in samples were distinguished on the basis of criteria that will be defined later (cf. 4.1.).

3.3. Plant biomarkers

Compounds supposedly likely to be recorded and to be discriminant found in soil samples and detected in plants are then described plant by plant. These compounds are listed in Table 2, together with other compounds that are *a priori* considered less specific.

3.3.1. *Sphagnum rubellum*

n-Alkanes were distributed from *n*-C₁₇ to *n*-C₃₃, with maxima at *n*-C₂₅, *n*-C₂₇ and *n*-C₂₉, and a strong odd predominance. *n*-Alkanols and *n*-alkanoic acids ranged from *n*-C₁₆OH to *n*-C₃₂OH and from *n*-C_{12:0} to *n*-C_{34:0}, respectively, exhibiting a strong even predominance. *n*-Alkanols and *n*-alkanoic acids had maximum values at *n*-C₂₈OH and at *n*-C_{16:0}, respectively. β - and α -amyrins and lupeol (tp2, tp5 and tp6), as well as their ketone counterparts (k2, k5

and k8, respectively) were the major pentacyclic triterpenes in the neutral fraction of *S. rubellum*. Some pentacyclic triterpenes were detected in trace amounts such as ketones (k3 and k11), triterpenol (tp3) and triterpenyl acetates (ta1, ta4, ta2 and ta7).

3.3.2. *Erica tetralix*

n-Alkanes ranged from *n*-C₂₁ to *n*-C₃₃ with maxima at *n*-C₂₇ and *n*-C₂₉ and a strong odd predominance. *n*-Alkanols and *n*-alkanoic acids ranged from *n*-C₁₁OH to *n*-C₃₂OH and from *n*-C_{10:0} to *n*-C_{34:0}, respectively. Both exhibited a strong even predominance and maximum at *n*-C₂₈OH, and at *n*-C_{16:0} and *n*-C_{28:0}, respectively. α -Amyrenyl acetate (ta4), several triterpene ketones (k2, k3, k5, k6, k8 and k12) and triterpenols (tp2, tp3, tp5 and tp6) were detected.

3.3.3. *Calluna vulgaris*

n-Alkanes were distributed from *n*-C₁₇ to *n*-C₃₅ with maxima at *n*-C₃₁ and *n*-C₃₃ and a strong odd predominance. *n*-Alkanols ranged from *n*-C₁₁OH to *n*-C₃₂OH with maximum at *n*-C₂₄OH and *n*-C₂₈OH, whereas *n*-alkanoic acids ranged from *n*-C_{10:0} to *n*-C_{34:0} with maximum at *n*-C_{16:0}. Both exhibited a strong even predominance. Some pentacyclic triterpenols (tp1, tp2, tp3, tp4, tp6 and tp8) and pentacyclic triterpene ketones (k1, k2, k4, k5, k7, k8 and k11) were detected, together with a wide diversity of pentacyclic triterpenyl acetates (ta1, ta4, ta2, ta3, ta5 and ta7).

3.3.4. *Molinia caerulea*

n-Alkanes were distributed from *n*-C₂₁ to *n*-C₃₃ with maxima at *n*-C₂₇ and *n*-C₂₉ and a strong odd predominance. *n*-Alkanols and *n*-alkanoic acids ranged from *n*-C₁₈OH to *n*-C₃₂OH and from *n*-C_{12:0} to *n*-C_{32:0}, respectively, exhibiting a strong even predominance. *n*-Alkanols

and *n*-alkanoic acids had maximum at *n*-C₂₈OH and at *n*-C_{16:0}, respectively. Except for *n*-alkanes, *n*-alkanoic acids and steroids, the major lipids in *M. caerulea* were essentially limited to β - and α -amyrenones (k2 and k5, respectively), lup-20(29)-en-3-one (lupenone; k12), fernenone (k10), β - and α -amyrins and lupeol (tp2, tp5 and tp6).

3.3.5. *Betula pendula*

n-Alkanes were distributed from *n*-C₁₅ to *n*-C₃₃ with maxima at *n*-C₂₅ and *n*-C₂₇ and a strong odd predominance. *n*-Alkanols and *n*-alkanoic acids ranged from *n*-C₁₂OH to *n*-C₃₂OH and from *n*-C_{10:0} to *n*-C_{34:0}, respectively, exhibiting a strong even predominance. *n*-Alkanols and *n*-alkanoic acids had maximum at *n*-C₂₈OH and at *n*-C_{16:0}, respectively. Pentacyclic triterpenes essentially consisted of lupane-derived compounds in both leaves and bark of *B. pendula*: lup-22(29)-en-3-one (k8), lupeol (tp6), lupa-2,20(29)-dien-28-ol (db2), betulin (db4) and betulinic acid (db2, Table 2). Methyl-3 α -acetoxylup-22(29)-en-28-oate (db1) was also detected in *B. pendula* bark.

3.3.6. *Pinus sylvestris*

n-Alkanes were distributed from *n*-C₁₅ to *n*-C₃₃ with maxima at *n*-C₂₅ and a strong odd predominance. *n*-Alkanols and *n*-alkanoic acids ranged from *n*-C₁₂OH to *n*-C₃₂OH and from *n*-C_{10:0} to *n*-C_{30:0}, with maxima at *n*-C₂₈OH and at *n*-C_{16:0}, respectively, and a strong even predominance. The needles and bark of *P. sylvestris* contained sterols and tricyclic diterpenes as major compounds. Tricyclic diterpenes were exclusively detected in *P. sylvestris*. Among several tricyclic diterpenes in *P. sylvestris*, only six were assigned (td1-6, Table 2). Five methoxy-serratenes (3 β ,21 α -dimethoxy-serrat-14-ene, ms1; 3 α -methoxy-serrat-14-en-21-one, ms2; 3 β -methoxy-serrat-14-en-21-one, ms3; 3-Methoxy-serrat-14-en-21-yl acetate, ms4 and

3-methoxy-serrat-14-en-21 α -ol, ms5) were found in notable concentrations in bark. In needles, only 3 β -methoxy-serrat-14-en-21-one (ms3) was detected and constituted the only pentacyclic triterpenes detected in *P. sylvestris*.

4. Discussion

4.1. Specificity of compounds in soil samples

With the exception of simiarenol, simiarenone and taraxasteryl acetate (tp7, k9 and ta6, respectively), all lipids detected in soil samples were also detected in plants (Table 2).-These compounds could thus be synthesized by others plants present in low abundance on La Guette peatland.

The low variations in concentration, distribution and ratios of *n*-alkyl lipids (*n*-alkanes, *n*-alkanols and *n*-alkanoic acids, Table 3) in soil samples, make these compounds inefficient for discriminating the surrounding vegetation. Several studies propose that *n*-alkyl lipids-based ratios can be used to reconstruct past vegetation dynamics in peatlands (Xie et al., 2004; Zhou et al., 2005; Nichols et al., 2006; Zheng et al., 2007; Zhou et al., 2010). For example, the *n*-C₂₃/*n*-C₂₉ (or *n*-C₂₃/*n*-C₃₁) *n*-alkanes ratio was used to estimate the relative input of *Sphagnum* species compared to vascular plants, based on the high proportions of *n*-C₂₃ in *Sphagnum* species. Similarly, and based on preliminary studies that estimated the relative proportions of submerged vs emerged and terrestrial plant in lacustrine systems (Ficken et al., 2000), the P_{aq} ratio [$P_{aq} = (n-C_{23} + n-C_{25}) / (n-C_{23} + n-C_{25} + n-C_{29} + n-C_{31})$] was proposed to estimate past hydrological conditions, based on the same observation that *Sphagnum* species mainly produce *n*-C₂₃ whereas vascular plants produce longer chain *n*-alkanes.

These parameters, although largely used in paleoenvironmental studies based on peat deposits, have been recently debated. Several *Sphagnum* species were shown to produce mainly $n\text{-C}_{27}$ or $n\text{-C}_{29}$ and not $n\text{-C}_{23}$ (Baas et al., 2000; Xie et al., 2004). Considering that *Betula* sp. produce also $n\text{-C}_{27}$ as the major n -alkane homologue, Andersson et al. (2011) concluded that the respective contribution of *Betula* sp. and *Sphagnum* sp. cannot be discriminated in peat deposits on the basis of P_{aq} and $n\text{-C}_{23}/n\text{-C}_{29}$. Additionally, Ronkainen et al. (2013) found similarities in $n\text{-C}_{23}/n\text{-C}_{29}$ n -alkane between *Sphagnum* spp. and vascular plants collected in bog peatlands. The efficiency of n -alkane ratios to differentiate species or plant group contributions in peatland is thus jeopardized.

$n\text{-C}_{27}$ is the dominant n -alkane in *S. rubellum* collected from La Gnette peatland, in agreement with Baas et al. (2000). This can explain the inefficiency of P_{aq} and $n\text{-C}_{23}/n\text{-C}_{29}$ in discriminating samples, considering also that there is no major difference in n -alkane distribution for vascular plants. In addition, it is worthwhile noting that our samples were collected on *M. caerulea* tussocks on which *Sphagnum* can barely develop. More generally, and according to Otto et al. (2005), it is difficult to classify vegetation types on the basis of n -alkyl lipid distributions considering the significant variability observed within the same plant family, genera or even species (e.g. Herbin and Robbins, 1969; Gülz et al., 1989).

Likewise, as steroids are present in all the analyzed plants, their distribution in soil samples was not discriminant. Derivatives of cholesterol are present in all eukaryote cells. To a large extent, C_{29} steroids derive mainly from the degradation of β -sitosterol (st4, highest content) and stigmasterol (st3) that originate from various plant sources (Harwood and Russell, 1984; Bianchi, 1995; Huang et al., 1995). The non-discriminative character of steroids in our samples is also in agreement with observations by Ronkainen et al. (2013) who indicated that sterols in soils are of limited interest for describing the surrounding vegetation.

Similarly, compounds with an oleanane or ursane skeleton and a Δ^{12} double bond such as pentacyclic triterpene acids Oa and Ua, as β - and α -amyryns, or as their ketone counterparts were not efficient for discriminating the surrounding vegetation. These compounds are synthesized by almost all plants sampled in the peatland, except *B. pendula* and *P. sylvestris*.

Pentacyclic triterpenyl acetates were detected only in Ericaceae species (*C. vulgaris* and *E. tetralix*), while pentacyclic triterpene ketones were overrepresented (Table 2) in these species, which are part of open vegetation. β -Amyrenyl, multiflorenyl, glutinyl, α -amyrenyl, isobauerenyl and friedelanyl acetates were found in *C. vulgaris*, whereas α -amyrenyl acetate (ta4) was the only pentacyclic triterpenyl acetate detected in *E. tetralix* (Table 2). Triterpenyl acetates have been proposed as specific markers for Asteraceae, but can also occur in other taxa (Lavrieux et al., 2011). Triterpenyl acetates found in high amount in the soil samples (ta2 and ta7) are those produced by typical peatland vegetation (*C. vulgaris*). Taraxasteryl acetate (ta6) was detected in low amounts except in StI and StIX. It was not found in any of the plants analyzed, but it has been described in several taxa such as Apocynaceae (Lavrieux et al., 2011 and references therein). According to the plant inventory of La Gnette carried out in 1999, 2002, 2011 and 2012 (Sologne Nature Environnement, 2012), only one Apocynaceae species was identified in high frequency at the site: *Gentiana pneumonanthe*. Further analysis should be done to ascertain whether this species contains pentacyclic triterpenyl acetates, and more specifically taraxasteryl acetate (ta6), as well as for compounds for which the source could not be identified (simiarenol and simiarenone). For Ericaceae, the distinction between *C. vulgaris* and *E. tetralix* could also be achieved by examining taraxerone, taraxerol, multiflorenone, glutinone, glutinol, friedelin and friedelanol (k1, tp1, k4, k7, tp4, k11 and tp8), which were only present in *C. vulgaris*, whereas isobauerenone (k3), hop-22(29)-en-3-one (k6) and lupenone (k12) were present in *E. tetralix*. Pancost et al. (2002) also observed

taraxerol (tp1) and taraxerone (k1) in *C. vulgaris*, whereas they were absent from *E. tetralix*. Detection of friedelin, friedelanol and friedelanyl acetate in our samples is in accordance with van-Smeerdijk and Boon (1987) who detected friedelane derivatives in Ericaceae species. Four methoxy-serratenes (ms1-3 and ms5) already described in the needles and/or bark of *P. sylvestris* by Norin and Winell (1972) were found in our soil samples and in *P. sylvestris*. The present study is the first report of the occurrence of such compounds in peatland soils, although their definition as specific biomarkers of Pinaceae has been demonstrated for mineral soils (Le Milbeau et al., 2013a). In addition, we detected 3,21-dimethoxy-serrat-14-en-29-yl acetate (ms4), also assigned to Pinaceae (Le Milbeau et al., 2013b). Six tricyclic diterpenes found in soil samples were also present in needles and bark of *P. sylvestris*. Tricyclic diterpenes are reputed specific biomarkers for conifers (i.e. Stefanova et al., 1995). Therefore, and considering their absence in other analyzed plants, both these compounds can be used as specific biomarkers of *P. sylvestris*. Tricyclic diterpenes concentrations were lower than those of methoxy-serratenes in all soil samples (Supplementary Data). The high correlation between the sum of tricyclic diterpenes concentrations and that of methoxy-serratenes ($r = 0.997$; $p \leq 0.0001$, $n=9$) confirms that they derive from the same source and did not suffer differential degradation.

Betulin, betulinic acid and their derivatives found in soil samples were also abundant in the extracts of leaves and bark from *B. pendula*. This is consistent with the report of betulinic acid, betulin and methyl-3 β -acetoxy-lup-22(29)-en-28-oate (db1, db3 and db4, respectively) in *Betula spp* bark by Schnell et al. (2013). In *B. pendula* samples (leaves and bark), betulin was overexpressed compared to lupeol (tp6). According to Ekman (1983) betulin extracted from the bark of Betulaceae represents up to 30% of dry wt. Ukkonen and Era (1979) showed that lupeol concentration decreased whereas betulin levels increased with tree age. In La

Guette peatland, lupeol cannot be solely associated to *B. pendula* because, although in low amounts, *M. caerulea* and *S. rubellum* also lupeol.

Lipids detected in *S. rubellum* corroborate the results of Baas et al. (2000) and Pancost et al. (2002). We identified steroids (st1-5), *n*-alkanes, *n*-alkanols, *n*-alkanoic acids, β -amyrin, α -amyrin, oleanoic acid, ursolic acid and lupeol (tp2, tp5, tp6, Oa and Ua, respectively). These compounds, as discussed below, were not specific. Additional terpenes (ta1, ta2, ta4, ta7, k3, k11, tp3, Table 2) detected in trace amounts have never been reported so far in *Sphagnum* species. Because these compounds were also detected in *C. vulgaris* and *E. tetralix*, a contamination of *S. rubellum* by these overlying plants cannot be excluded.

The analysis of soil samples combined with that of the main plants growing in La Guette peatland thus allowed us discriminating between non-specific and specific compounds, i.e. those that potentially allow depicting the surrounding vegetation from the chemical imprints of soils. The specific character of biomarkers can rely on several criteria. In our case, the specificity of tricyclic diterpenes was linked to their skeleton, as for betulin derivatives, except for lupeol and lupenone. The specificity of methoxy-serratenes relies both on their original skeleton and on the presence of a methoxyl group. For triterpenyl acetates, their specificity is due to the acetate functional group, independently of the pentacyclic triterpene skeleton. Early diagenetic processes occurring in soils could alter the specificity of compounds to its source plant. For example, Rullkötter et al. (1994) and ten Haven et al. (1992) showed that pentacyclic triterpenes can suffer structural rearrangements and functional group transformation. If this can account for biomarkers that were shown not to be specific (amyrins and amyrenones for example), this had no or little influence on the biomarkers listed above as specific of La Guette vegetation. Although early diagenesis can affect tricyclic diterpenes, it will produce other tricyclic diterpenes that will remain specific for *P. sylvestris*.

The same applies for methoxy-serratenes and, to a lesser extent, to betulin derivatives. Therefore, we considered that early diagenesis [that is mild considering high CPI values of *n*-alkyl lipids (Table 3), and the absence of degraded triterpenes] did not alter the plant-biomarker relationships established here above.

On this basis, we defined (i) k1, k4, k7, k11, tp1, tp4, tp8, ta2, ta3, ta5 and ta7 as tracers of *C. vulgaris*, (ii) k3 and k12 as tracers of *E. tetralix*, (iii) methoxy-serratenes (ms1-5) and tricyclic diterpenes (td1-6) as tracers of *P. sylvestris*, and (iv) db1, db2, db3 and db4 as tracers of *B. pendula*.

4.2. Distribution of individual compounds in open vs. closed stations

To discriminate the sampling stations based on the presence/absence and on the abundance of the biomarkers in soil samples, a PCA was undertaken. Ericaceous biomarkers have high loadings on Factor 1 (45.95% of the variance, compounds k1, k3, k4, k7, k12, tp1, tp4, tp8, ta2, ta3, ta5 and ta7 all positive loadings; Fig. 3a) and to a less extent on Factor 3 (9% of the variance, k11 negative loading; Fig. 3a). These two axes are defined by Ericaceae biomarkers and thus may discriminate sites based on *C. vulgaris* and *E. tetralix* abundances: StII and StIII on one side and the other sites on the other side (Fig. 3b). Because we could not estimate the biomass of these two species we cannot further interpret the distribution of samples and variables in the F1/F3. Nevertheless, these results suggest that there might be a control of Ericaceae biomarkers for which much attention should be paid in future research.

We then focus on Factor 2 and Factor 4 that allow discriminating between biomarkers of *P. sylvestris* and *B. pendula*, for which we have quantitative biomass data. *P. sylvestris* biomarkers have high loadings on Factor 2 (35.68% of the variance, compounds ms1, ms3, ms4, ms5, td1, td2, td3, td4, td5 and td6; Fig. 3c). Stations StV and StIII, the two stations with

the highest *P. sylvestris* biomass index (BI_P; Table 1), had the highest loading on the Factor 2 axis (Fig. 3d). This shows that larger amounts of *P. sylvestris* specific biomarkers were found in the soil when this plant is abundant in the surroundings.

B. pendula biomarkers db1 and db3 have high loadings on Factor 4 (5.23% of the variance; Fig. 3c). StVIII, which displays the highest biomass index of *B. pendula* (BI_B; Table 1), had the highest loading on the Factor 4 axis (Fig. 3b) that corresponds to *B. pendula* specific biomarkers. The stations others than StV and StIII (*P. sylvestris*) and StVIII (*B. pendula*) are the most open stations, with total biomass index (BI_T; Table 1) lower than 0.50. The best relationships were obtained for abundant compounds such as db3, ms1, ms3, ms4, ms5 and tricyclic diterpenes, whereas some discrepancies are noted for the less abundant compounds (db1, db2, db4 and ms2). Thus, and except for these minor compounds, PCA showed that overlying tree vegetation clearly imprints its chemical signature on underlying soils. Three different vegetation cover could be distinguished: (i) closed vegetation dominated by *P. sylvestris*; (ii) closed vegetation dominated *B. pendula* and (iii) open vegetation (StI and StII) or semi-open vegetation (early colonization by trees, StIV, StVI, StVII, StIX).

4.3. Influence of tree biomass on soil biomarker concentration

We then further explore whether biomarker concentrations can be used to quantitatively reflect tree biomass in the neighborhood of sampling sites. For this purpose, the individual and combined concentrations of specific biomarkers of *P. sylvestris* (only methoxy-serratenes since methoxy-serratenes and tricyclic diterpenes concentrations are correlated, cf. 4.1.) and of *B. pendula* (betulin derivatives: db1-4; Supplementary data) were compared to the respective biomass index (BI_P and BI_B), estimated at each station (n=9; Fig. 4a). This allowed determining a correlation (R) between biomarker concentration in soils and BI for each area.

Figure 4a gives an example for total methoxy-serratenes where, for a surface of 12.8 m², R was 0.94.

Then, R values obtained are displayed against increasing area for total and individual *P. sylvestris* biomarkers (Fig. 4b) and for total and individual *B. pendula* biomarkers (Fig. 4c). In the smallest area (3.16 m², Fig. 4bc), the correlations equaled zero as there was no trees just at the point of sampling. For *P. sylvestris*, there was a good correlation between methoxy-serratenes other than ms2 and the BI_P (Y axis on Fig. 4a, R = 0.45). Inversely, for *B. pendula*, there was an only one good correlation between db3 and the BI_B (Fig. 4b, R = 0.91; P<0.001). The lowest correlations were observed for db1, db2, db4 and ms2, i.e. biomarkers detected in low amounts. The best correlations were obtained from 28.3 m² for the total concentration of methoxy-serratenes (R > 0.96, P<0.001) and for ms1, ms3 and ms4, after an intermediate state where lower correlations were found (Fig. 4b). The best correlations were obtained from 12.6 m² for the total concentration of betulin derivatives (R = 0.91, P<0.001) as well as for db3 (Fig. 4c). In both cases, the quantitative relationships between soil biomarker concentrations and tree biomass are valid up to, at least, 80 m². These results show that biomarker concentrations at a specific site are not solely representative of vegetation biomass present directly above the sampling point, but integrate a larger surface. The use of selected biomarkers for these relationships could limit any bias related to individual biomarker behavior (variations in production, transport, or early diagenesis).

4.4. Potential in palaeoenvironmental studies

Peatlands are prone to successional change since plant growth causes the filling in of lakes and the accumulation of peat on waterlogged sites (Hogg et al., 1995) and the OM accumulated over time registers the compositional changes in vegetation (Comont et al.,

2006; Gogo et al., 2011). These ecosystems can develop from aquatic systems to forested areas (terrestrialization). At any developmental stage, events such as fire or hydrological modifications can (i) reset the system to previous stage, or (ii) accelerate the natural dynamics. As such, reopening and reclosing of the system can occur. Indeed, in La Guette peatland, tree cover increased from the 1940's to the mid 1970's. In 1976, a fire eliminated all the trees and opened the site.

The relationships between biomarkers retrieved from sedimentary archives and their source organisms has been used for decades to reconstruct paleovegetations (i.e. Cranwell, 1984; Volkman, 1986; Rieley et al., 1991; Meyers, 2003; Xie et al., 2004; Jacob et al., 2005; Jia et al., 2008; Zocatelli et al., 2010; Regenery et al., 2013). However, and as noted by Eglinton and Eglinton (2008), there are remaining questions on (i) the spatial relationships embedded in sedimentary biomarkers and on (ii) the processes that control their input, notably through production, dispersal, and sedimentation. For instance, Stefanova et al. (2011) and Regnery et al. (2013) have shown that some biomarkers archived in sediments were in lower abundance than expected on the basis of pollen data, highlighting the difficulties in interpreting the meaning and the representativeness of biomarker concentrations compared to pollen data.

Our innovative approach allows us determining the spatial and quantitative imprint of present-day vegetation on underlying soils. Such an approach should be expanded in order to better constrain the quantitative representativeness of biomarkers in sedimentary archives such as peats and the spatial heterogeneities of ancient peatlands. Tiny discrepancies in our data could result from differential diagenesis (although we expect similar behavior for compounds of similar structures) or could be the expression of the contribution of past

vegetation biomarkers to the soil that would bias/blur the direct relationship between vegetation distribution and soil biomarkers.

5. Conclusions

This study is one of the first attempts to explore spatial chemodiversity in soils related to biodiversity based on an exhaustive biomarker inventory in peatland soils compared to molecular imprints of surrounding vegetation and then to the evaluation of tree biomass. Among the relevant specific biomarkers tricyclic diterpenes and methoxy-serratenes stand out as specific biomarkers of *P. sylvestris*; and betulin derivatives as specific biomarkers of *B. pendula*, as already defined in previous studies. Here, Taraxerone, Multiflorenone, Glutinone, Friedelin, Taraxerol, Glutinol, Friedelanol, Multiflorenyl acetate, Glutinyll acetate, Isobauerenyl and Friedelanyl acetate were defined as tracers of *C. vulgaris* and k3 and k12 as tracers of *E. tetralix*.

Multivariate analysis of soil biomarker concentrations reveals that they describe the overlying vegetation, allowing distinguishing three different vegetation cover: (i) closed vegetation dominated by *P. sylvestris*; (ii) closed vegetation dominated *B. pendula* and (iii) open vegetation or semi-open vegetation (early colonization by trees). Quantitative information on tree biomass could be estimated through specific biomarker concentration. This study contributes to our understanding of how biomarker distribution and concentration in spatially constrained archives, such as peat cores, can render the spatial heterogeneity of vegetation distribution that formerly covered the catchment. Such an approach could allow paleo-reconstructions to better account for the diversity of contexts in palaeolandscapes.

Acknowledgements

This work was part of the PRINCIPASOL project (<http://www-projets.isto.cnrs-orleans.fr/principasol/>) supported by the Conseil Général du Loiret and FEDER funding. S.G. was funded by the Spatioflux project and the Labex VOLTAIRE (ANR-10-LABX-100-01). We are grateful to J. Sauze for the GIS map layers used for Fig. 1. We acknowledge those who helped during sampling, in particular M.-P. Gauthier (Aromatechnologies) and B. Soengas (LVMH-Recherche). The study site, La Guette peatland, is part of the CNRS Service National d'Observation Tourbières. The authors also wish to thank the anonymous reviewers their constructive comments.

References

- Aerts, R., Wallen, B., Maimor, N., 1992. Growth-limiting nutrients Sphagnum-dominated bogs subject to low and high atmospheric N supply. *Journal of Ecology* 80, 131-40.
- Andersson, R.A., Kuhry, P., Meyers, P., Zebühr, Y., Crill, P., Mörtz, M., 2011. Impacts of paleohydrological changes on *n*-alkane biomarker compositions of a Holocene peat sequence in the eastern European Russian Arctic. *Organic Geochemistry* 42, 1065-1075.
- Baas, M., Pancost, R.D., van, Geel, B., Sinninghe Damsté, J.S., 2000. A comparative study of lipids in Sphagnum species. *Organic Geochemistry* 31, 535-541.
- Berendse, F., Schmitz, M., Isser, V., 1994. Experimental manipulation of succession in heathland ecosystems. *Oecologia* 100, 38-44.
- Bianchi, G., 1995. Plant waxes. In: Hamilton, R.J. (Ed.), *Waxes: Chemistry, molecular biology and functions*. The Oily Press, Dundee, 175-222 pp.

- Binet, S., Gogo, S., Laggoun-Défarge, F., 2013. A water-table dependent reservoir model to investigate the effect of drought and vascular plant invasion on peatland hydrology. *Journal of Hydrology* 499, 132-139.
- Bobbink, R., Lamers, L.P.M., 2002. Effect of increased nitrogen deposition. *Air Pollution and Plant Life*, 2nd edn (eds J.N.B. Bell & M. Treshow), John Wiley & Sons, Ltd., New York, 201-235 pp.
- Boeye, D., Vehagen, B., Van Haesebroeck, V., Mohssine, E., 1999. Phosphorus fertilization in a phosphorus-limited fen: effects of timing. *Applied Vegetation Science* 2, 71-78.
- Chambers, F. M., Mauquoy, D., Todd, P.A., 1999. Recent rise to dominance of *Molinia caerulea* in environmentally sensitive areas: new perspectives from palaeoecological data. *Journal of Applied Ecology* 36, 719-733.
- Comont, L., Laggoun-Défarge, F., Disnar, J.R., 2006. Evolution of organic matter indicators in response to major environmental changes: The case of a formerly cut-over peat bog (Le Russey, Jura Mountains, France). *Organic Geochemistry* 37, 1736-1751.
- Cranwell, P.A., 1984. Organic geochemistry of lacustrine sediments: triterpenoids of higher plant origin reflecting post-glacial vegetational succession. In "Lakes sediments and environmental History" (Edited by Haworth E.Y. and Lund J.W.G.). Univ. Press, Leicester, U.K, 69-92 pp.
- Eglinton, T.I., Eglinton, G., 2008. Molecular proxies for paleoclimatology. *Earth and Planetary Science Letters* 275, 1-16.
- Ekman, R., 1983. The suberin monomers and triterpenoids from the outer bark of *Betula verrucosa* Ehrh. *Holzforschung* 37, 205-211.

- Ficken, K.J., Li, B., Swain, D.L., Eglinton, G., 2000. An n-alkane proxy for the sedimentary input of submerged/floating freshwater aquatic macrophytes. *Organic Geochemistry* 31, 745–749.
- Gimingham, C., 1972. *Ecology of heathlands*. Chapman and Hall, London, UK.
- Gogo, S., Laggoun-Défarge, F., Delarue, F., Lottier, N., 2011. Invasion of a *Sphagnum*-peatland by *Betula* spp and *Molinia caerulea* impacts on organic matter biochemistry. Implications for carbon and nutrient cycling. *Biogeochemistry* 106, 53-69.
- González-Pérez, J.A., González-Vila, F.J., Arias, M.E., Rodríguez, J., Marañón, T., Clemente, L., 2011. Geochemical and ecological significance of soil lipids under *Rhododendron ponticum* stands. *Environmental Chemistry Letters* 9, 453-466.
- Gülz, P.-G., Müller, E., Prasad, R.B.N., 1989. Organ-specific composition of epicuticular waxes of beech (*Fagus silvatica* L.) leaves and seeds. *Zeitschrift für Naturforschung* 44, 731-734.
- Hansen, K., 1976. Ecological studies in Danish heath vegetation. *Dansk Botanisk Archiv* 31, 7-116.
- Harwood, J.L., Russell, N.J., 1984. *Lipids in Plants and Microbes*. George Allen and Unwin, London, 162 pp.
- Herbin, G.A., Robbins, P.A., 1969. Patterns of variation and development in leaf wax alkanes. *Phytochemistry* 8, 1985-1998.
- Hogg, P., Squires, P., Fitter, A.H., 1995. Acidification, nitrogen deposition and rapid vegetational change in a small valley mire in Yorkshire. *Biological Conservation* 71, 143-153.

- Huang, Y., Lockheart, M.J., Collister, J.W., Eglinton, G., 1995. Molecular and isotopic biogeochemistry of the Miocene Clarkia Formation: hydrocarbons and alcohols. *Organic Geochemistry* 23, 785-801.
- Hunt, R., Hand, D.W., Hannah, M.A., Neal, A.M., 1991. Response to CO₂ enrichment in 27 herbaceous species. *Functional Ecology* 5, 410-421.
- Hunt, R., Hand, D.W., Hannah, M.A., Neal, A.M., 1993. Further responses to CO₂ enrichment in British herbaceous species. *Functional Ecology* 7, 661-668.
- Jacob, J., Disnar, J.R., Boussafir, M., Albuquerque, A.L.S., Sideffine, A., Turcq, B., 2005. Pentacyclic triterpene methyl ethers in recent lacustrine sediments (Lagoa do Caçó, Brazil). *Organic Geochemistry* 36, 449-461.
- Jacquemyn, H., Brys, R., Neubert, M.G., 2005. Fire increases invasive spread of *Molinia caerulea* mainly through changes in demographic parameters. *Ecological Applications* 15, 2097-2108.
- Jia, G., Dungait, J.A.J., Bingham, E.M., Valiranta, M., Korhola, A., Evershed, R.P., 2008. Neutral monosaccharides as biomarker proxies for bog-forming plants for application to palaeovegetation reconstruction in ombrotrophic peat deposits. *Organic Geochemistry* 39, 1790-1799.
- Lavrieux, M., Jacob, J., Le Milbeau, C., Zocatelli, R., Masuda, K., Bréheret, J.G., Disnar, J.R., 2011. Occurrence of triterpenyl acetates in soil and their potential as chemotaxonomical markers of Asteraceae. *Organic Geochemistry* 42, 1315-1323.
- Lavrieux, M.L., Bréheret, J.G., Disnar, J.R., Jacob, J., Le Milbeau, C., Zocatelli, R., 2012. Preservation of an ancient grassland biomarker signature in a forest soil from the French Massif Central. *Organic Geochemistry* 51, 1-10.

- Le Milbeau, C., Lavrieux, M., Jacob, J., Bréheret, J.G., Zocatelli, R., Disnar, J.R., 2013a. Methoxy-serratenes in a soil under conifers and their potential use as biomarkers of Pinaceae. *Organic Geochemistry* 55, 45-54.
- Le Milbeau, C., Jacob, J., Zocatelli, R., Gogo, S., Laggoun-Defarge, F., 2013b. Origin of dimethoxyserratene derivatives in a peat bog. Abstract. 26th International Meeting on Organic Geochemistry, Tenerife.
- Legendre, P., Legendre, L., 1998. Numerical ecology. Developments in environmental modelling 20. Second English edition. Netherlands, Elsevier Science B.V., 853 pp.
- Limpens, J., Berendse, F., Blodau, C., Canadell, J.G., Freeman, C., Holden, J., Roulet, J., Rydin, H., Schaepman-Strub, G., 2008. Peatlands and the carbon cycle: from local processes to global implications – a synthesis. *Biogeosciences* 5, 1475-1491.
- Marseille, F., Disnar, J.R., Guillet, B., Noack, Y., 1999. *n*-Alkanes and free fatty acids in humus and A1 horizons of soils under beech, spruce and grass in the Massif Central (Mont-Lozère), France. *European Journal of Soil Science* 50, 433-441.
- McClymont, E.L., Bingham, E.M., Nott, C.J., Chambers, F.M., Pancost, R.D., Evershed, R.P., 2011. Pyrolysis GC–MS as a rapid screening tool for determination of peat-forming plant composition in cores from ombrotrophic peat. *Organic Geochemistry* 42, 1420-1435.
- Meyers, P.A., 2003. Applications of organic geochemistry to paleolimnological reconstructions: a summary of examples from the Laurentian Great Lakes. *Organic Geochemistry* 34, 261-289.
- Nichols, J.E., Booth, R.K., Jackson, S.T., Pendall, E.G., Huang, Y., 2006. Paleohydrologic reconstruction based on *n*-alkane distributions in ombrotrophic peat. *Organic Geochemistry* 37, 1505-1513.

- Norin, T., Winell, B., 1972. Extractives from the bark of scots pine, *Pinus sylvestris* L. Acta Chemica Scandinavica 26, 2297-2304.
- Otto, A., Simpson, M.J., 2005. Degradation and preservation of vascular plant derived biomarkers in grassland and forest soils from Western Canada. Biogeochemistry 74, 377-409.
- Pancost, R.D., Baas, M., van Geel, B., Damsté, J.S.S., 2002. Biomarkers as proxies for plant inputs to peats: an example from a sub-boreal ombrotrophic bog. Organic Geochemistry 33, 675-690.
- Regnery, J., Püttmann, W., Koutsodendris, A., Mulch, A., Pross, J., 2013. Comparison of the paleoclimatic significance of higher land plant biomarker concentrations and pollen data: A case study of lake sediments from the Holsteinian interglacial. Organic Geochemistry, 73-84.
- Rieley, G., Collier, R.J., Jones, D.M., Eglinton, G., 1991. The biogeochemistry of Ellesmere Lake, U.K. I: source correlation of leaf wax inputs to the sedimentary record. Organic Geochemistry 17, 901-912.
- Ronkainen, T., McClymont, E.L., Väiliranta, M., Tuittilac, E-S., 2013. The *n*-alkane and sterol composition of living fen plants as a potential tool for palaeoecological studies. Organic Geochemistry 59, 1-9.
- Rullkötter, J., Peakman, T.M., ten Haven, H.L., 1994. Early diagenesis of terrigenous triterpenoids and its implications for petroleum geochemistry. Organic Geochemistry 21, 215-233.
- van Smeerdijk, D., Boon, J.J., 1987. Characterisation of subfossil sphagnum leaves, rootlets of ericaceae and their peat by pyrolysis high-resolution gas chromatography-mass spectrometry. Journal of Analytical and Applied Pyrolysis 11, 377-402.

- Schnell, G., Schaeffer, P., Tardivon, H., Motsch, E., Connan, J., Ertlen, D., Schwartz, D., Schneider, N., Adam, P., 2013. Contrasting diagenetic pathways of higher plant triterpenoids in buried wood as a function of tree species. *Organic Geochemistry* 66, 107-124.
- Sologne Nature Environnement, SNE, 2012. Principasol: Inventaires floristiques et cartographies d'habitats. Bilan 2011-2012. Principasol. 31 pp.
- Statsoft Inc, 2008. STATISTICA for Windows Version 8.0. Tulsa Oklahoma, Statsoft, Inc.
- Stefanova, M., Ivanov, D.A., Utescher, T., 2011. Geochemical appraisal of palaeovegetation and climate oscillation in the Late Miocene of western Bulgaria. *Organic Geochemistry* 42, 1363-1374.
- Stefanova, M., Simoneit, B.R.T., Stojanova, G., Nosyrev, I., Goranova, M., 1995. Composition of the extract from a Carboniferous bituminous coal-I. Bulk and molecular constitution. *Fuel* 74, 768-778.
- ten Haven, H. L., Rullkötter, J., Stein, R., 1992. Preliminary analysis of extractable lipids in sediments from the eastern North Atlantic (leg 108): comparison of a coastal upwelling area (site 658) with a non upwelling area (site 659). *Proceedings of the Ocean Drilling Programme Scientific Results* 108, 351-360.
- Tomassen, H.B.M., Smolders, A.J.P., Limpens, J., Lamers, L.P.M., Roelofs, J.G.M., 2004. Expansion of invasive species on ombrotrophic bogs: desiccation or high N deposition? *Journal of Applied Ecology* 41, 139-150.
- Tomassen, H.B.M., Smolders, A.J.P., Lamers, L.P.M., Roelofs, J.G.M., 2003. Stimulated growth of *Betula pubescens* and *Molinia caerulea* on ombrotrophic bogs: role of high levels of atmospheric nitrogen deposition. *Journal of Ecology*, 91, 357-370.

- Trendel, J.M., Schaeffer, P., Adam, P., Ertlen, D., Schwartz, D., 2010. Molecular characterisation of soil surface horizons with different vegetation in the Vosges Massif (France). *Organic Geochemistry* 41, 1036-1039
- Ukkonen, K., Era, V., 1979. Birch bark extractives. *Kemia-Kemi* 6, 217-220.
- Volkman, J.K., 1986. A review of sterol markers for marine and terrigenous organic matter. *Organic Geochemistry* 9, 83-99.
- Woodin, S. J., Farmer, A.M., 1993. Impacts of sulphur and nitrogen deposition on sites and species of nature conservation importance in Great Britain. *Biological Conservation* 63, 23-30.
- Xie, S., Nott, C.J., Avsejs, L.A., Maddy, D., Chambers, F.M., Evershed, R.P., 2004. Molecular and isotopic stratigraphy in an ombrotrophic mire for paleoclimate reconstruction. *Geochimica et Cosmochimica Acta* 68, 2849-2862.
- Zheng, Y., Zhou, W., Meyers, P.A., Xie, S., 2007. Lipid biomarkers in the Zoigê-Hongyuan peat deposit: Indicators of Holocene climate changes in West China. *Organic Geochemistry* 38, 1927-1940.
- Zhou, W., Xie, S., Meyers, P.A., Zheng, Y., 2005. Reconstruction of late glacial and Holocene climate evolution in southern China from geolipids and pollen in the Dingnan peat sequence. *Organic Geochemistry* 36, 1272-1284.
- Zhou, W., Zheng, Y., Meyers, P.A., Jull, A.J.T., Xie, S., 2010. Postglacial climate-change record in biomarker lipid compositions of the Hani peat sequence, northeastern China. *Earth and Planetary Science Letters* 294, 37-46.
- Zocatelli, R., Jacob, J., Turcq, B., Boussafir, M., Sifeddine, A., Bernardes, M.C., 2010. Molecular evidence for recent turf cultivation in Northeast Brazil (Lagoa do Boqueirão, RN State). *Organic Geochemistry* 41, 427-430.

Zocatelli, R., Lavrieux, M., Disnar, J.-R., Le Milbeau, C., Jacob, J., Bréheret, J.G., 2012. Free fatty acids in Lake Aydat catchment soils (French Massif Central): sources, distributions and potential use as sediment biomarkers. *Journal of Soils and Sediments* 12, 734-748.

FIGURE CAPTIONS

Fig. 1. Location of La Guette peatland and sampling stations. Photos of four stations illustrate the vegetation cover depending on the degree of invasion by vascular plants. In the site map the areas partly invaded by *B. pendula* and *P. sylvestris* are shown in gray; the part covered by open vegetation (*Sphagnum*, *E. tetralix* and *C. vulgaris*) is shown in white.

Fig. 2. Ketone/acetate and alcohol fractions extracted from soil samples collected at different stations (StII, StIII, StV and StIX) illustrated in Figure 1. Nomenclature of compounds refers to Table 2.

Fig. 3. Variables (a) and individuals (b) PCA plots. In the variable plot, Factor 1 axis corresponds to betulinic acid and methyl-3 β -acetoxylup22(29)-en-28-oate (db3 and db1, respectively) originating from *B. pendula*, Factor 2 axis corresponds to biomarkers of *P. sylvestris* (ms1, ms2-5, td1-6) and Factor 3 axis corresponds to biomarkers of Ericaceae.

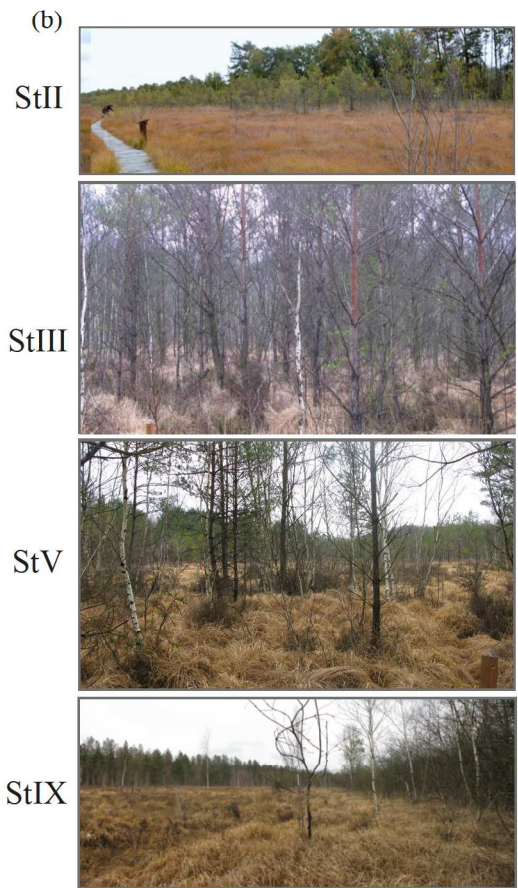
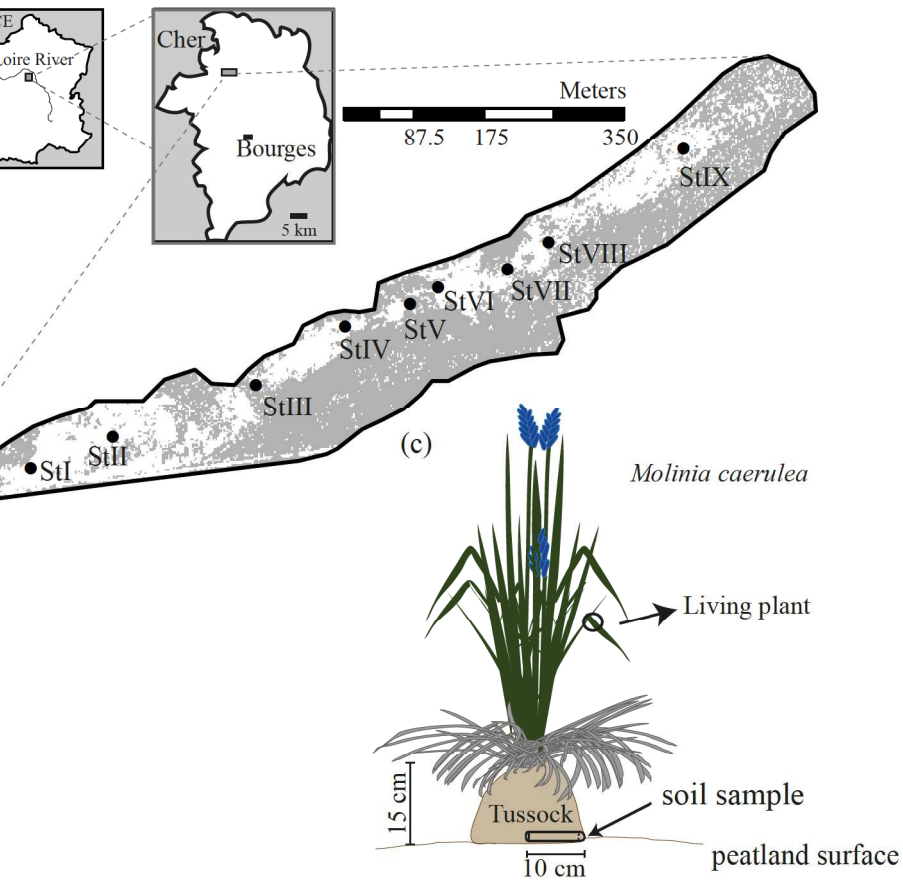
Fig. 4. Correlation between total methoxy-serratenes concentrations and biomass index (BI_P, calculated at 28.3 m²) of *P. sylvestris* (a), and the evolution of between these correlations with increasing area (from 3.1 m² to 78.5 m²) used to calculate the BI for individual methoxy-serratenes (b) and betulin derivatives (c) concentrations.

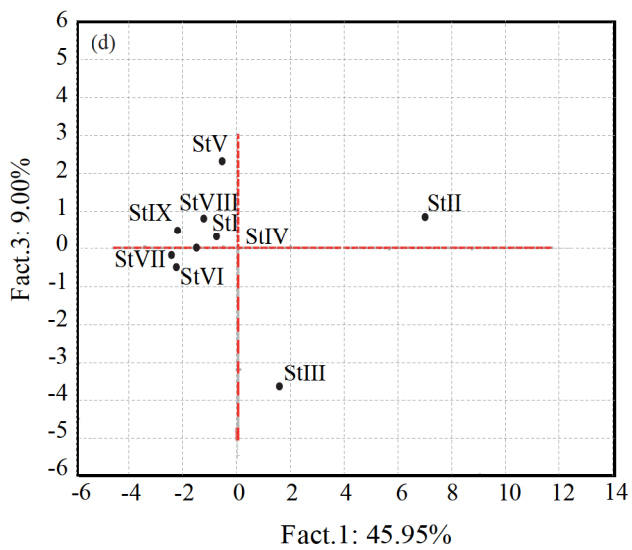
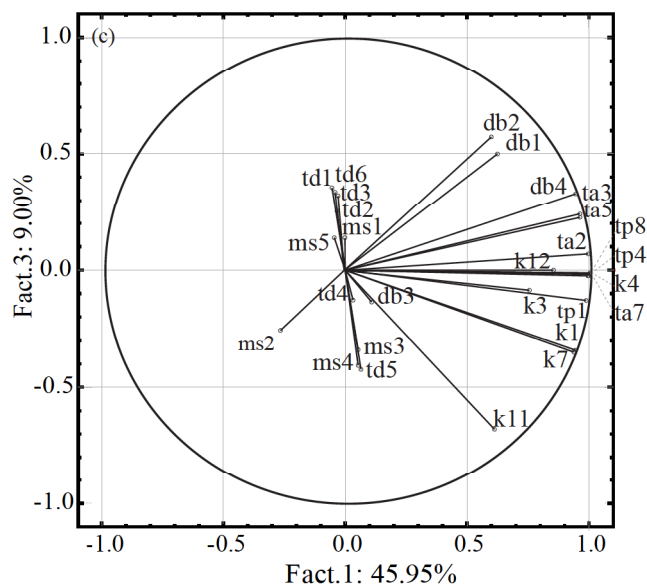
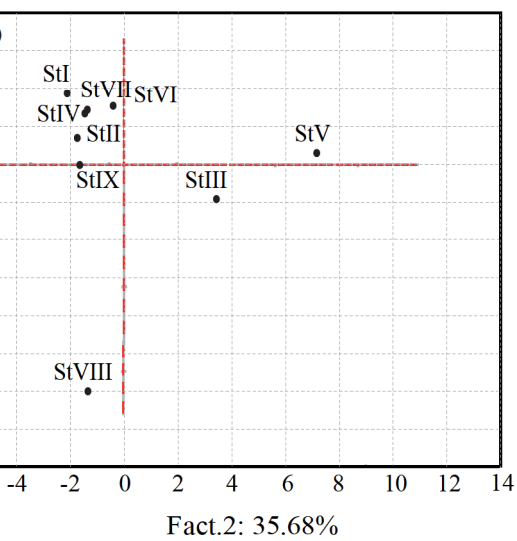
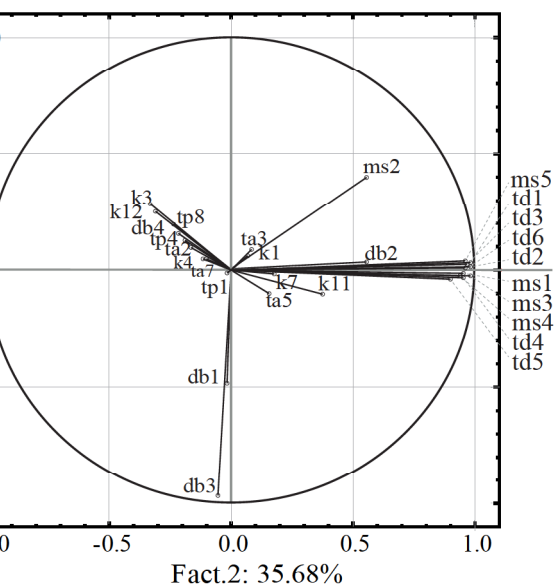
TABLE CAPTIONS

Table 1. Biomass index and percentage of *P. sylvestris* and *B. pendula* calculated in each station from tree counting and height measurements in La Gnette peatland, spring 2011.

Table 2. Presence and concentration of molecular biomarkers in surface soil and plant samples from La Gnette peatland. Numbers refer to peaks in the chromatograms displayed in Fig. 2: x, present; tr trace; - not detected.

Table 3. Concentrations and distribution parameters (ACL, Paq, CPI...) of *n*-alkyl lipids in soil samples.





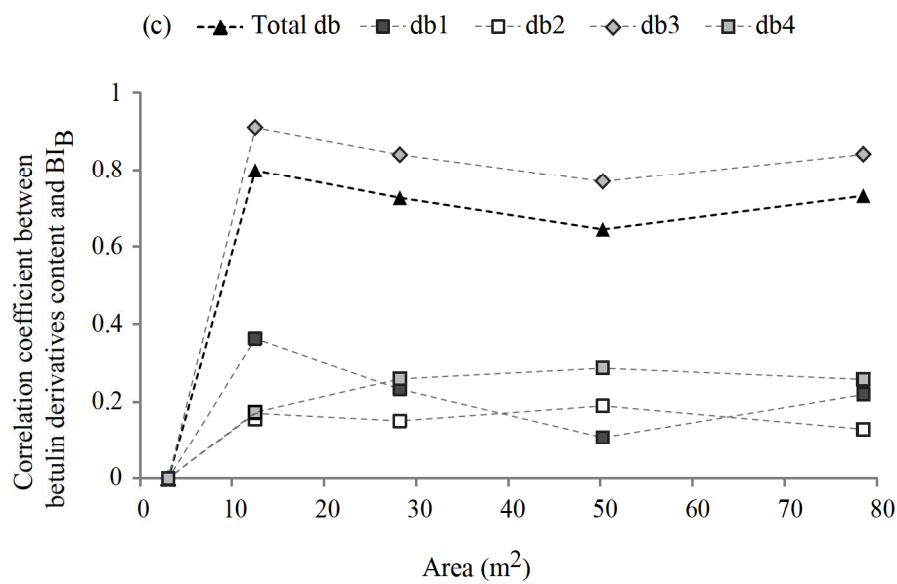
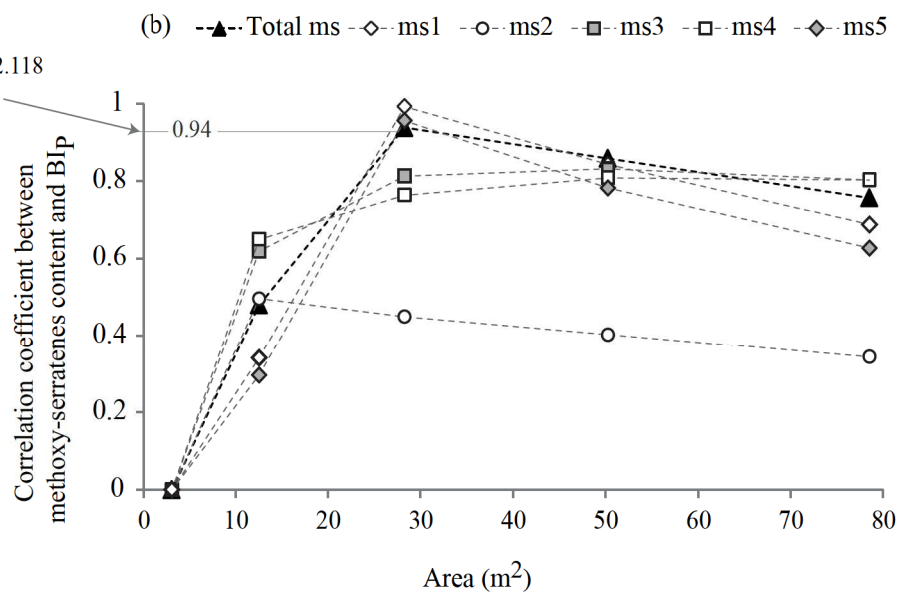
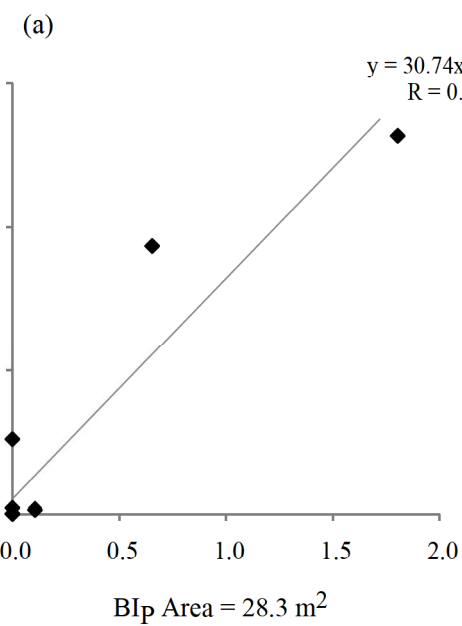


Table 1

Calculation from tree count (r - 5mt; a - 78.5 m²) in La Gvette peatland, spring 2011.

	Sampling points								
	StI	StII	StIII	StIV	StV	StVI	StVII	StVIII	StIX
BI _P ^a (tree/m ²)	0.00	0.00	1.09	0.00	0.70	0.00	0.20	0.70	0.00
BI _B ^b (tree/m ²)	0.00	0.00	0.43	0.04	0.20	0.01	0.30	0.60	0.17
BI _T ^c (tree/m ²)	0.00	0.00	1.53	0.04	0.90	0.01	0.50	1.30	0.17
P _% ^d (%)	-	-	72	0	78	0	37	51	0
B _% ^e (%)	-	-	28	100	22	100	63	49	100

^a Biomass index of *P. sylvestris*, ^b Biomass index of *B. pendula*, ^c Biomass index of trees, ^d percentage of *P. sylvestris*, ^e percentage of *Betula pendula*.

Table 2:

Presence of molecular biomarkers in surface soil and plant samples from La Guette peatland (Numbers refer to peaks in the chromatograms on Fig. 2: x, present; tr trace; - not detected).

#	Compound	Trivial name	Peat samples	<i>S. rubellum</i>	<i>E. teralix</i>	<i>C. vulgaris</i>	<i>M. caerulea</i>	<i>B. pendula</i> bark	<i>B. pendula</i> leaves	<i>P. sylvestris</i> bark	<i>P. sylvestris</i> needles
<i>n</i> -Alkyl lipids											
<i>n</i> -Alkanes											
			C _{15:35}	C _{17:33}	C _{21:33}	C _{17:35}	C _{21:33}	C _{15:33}	C _{15:33}	C _{15:33}	C _{15:33}
<i>n</i> -Alkanols											
			C _{19:31}	C _{16:32}	C _{12:32}	C _{12:32}	C _{16:32}	C _{12:32}	C _{12:32}	C _{12:31}	C _{12:31}
<i>n</i> -Alkanoic acids											
			C _{14:32}	C _{12:34}	C _{10:34}	C _{10:34}	C _{12:32}	C _{10:34}	C _{10:34}	C _{10:30}	C _{10:30}
Cyclic lipids											
Steroids											
st1	Cholest-5-en-3β-ol	Cholesterol	x	x	x	x	x	x	x	x	x
st2	24-Methyl-cholest-5-en-3β-ol	Campesterol	x	x	x	x	x	x	x	x	x
st3	24-Ethyl-cholesta-5,22-dien-3β-ol	Stigmasterol	x	x	x	x	x	x	x	x	x
st4	24-Ethyl-cholest-5-en-3β-ol	Sitosterol	x	x	x	x	x	x	x	x	x
st5	24-Ethyl-cholestan-3β-ol	Sitostanol	x	x	x	x	x	x	x	x	x
Tricyclic diterpenes											
td1	Podocarp-7-en-3-one, 13β-methyl, 13 vinyl	Pimara-7,15-dien-3-one	x	-	-	-	-	-	-	x	x
td2	Podocarp-8(14)-en-3-one		x	-	-	-	-	-	-	x	x
td3	Podocarp-8,11,13-trien-15al, 13 isopropyl	Abieta-8,11,13-trien-18-al	x	-	-	-	-	-	-	x	x
td4	Podocarp-7-en-15oic acid, 13β-methyl-13 vinyl methyl ester	Methyl isopimarate	x	-	-	-	-	-	-	x	x
td5	Podocarp-8,11,13-trien-15-oic acid,13 isopropyl methyl ester	Dehydroabietic acid, methyl ester	x	-	-	-	-	-	-	x	x
td6	Podocarpa-7,13-dien-15-oic acid, 13-isopropyl methyl ester	Methyl abietate	x	-	-	-	-	-	-	x	x
Methoxy-serratenes											
ms1	3β,21α-Dimethoxy-serrat-14-ene		x	-	-	-	-	-	-	x	-
ms2	3α-Methoxy-serrat-14-en-21-one		x	-	-	-	-	-	-	x	-
ms3	3β-Methoxy-serrat-14-en-21-one		x	-	-	-	-	-	-	x	x
ms4	3-Methoxy-serrat-14-en-21-yl acetate		x	-	-	-	-	-	-	x	-
ms5	3-Methoxy-serrat-14-en-21α-ol		x	-	-	-	-	-	-	x	-
Pentacyclic triterpenes											
Pentacyclic triterpenyl acetates											
ta1	Olean-12-en-3β-yl acetate	β-Amyrenyl acetate	x	tr	-	x	-	-	-	-	-

ta2	D:C-friedo-olean-7-en-3β-yl acetate	x	tr	-	x	-	-	-	-	-	-	-
ta3	Glutin-5-en-3b-yl acetate	x	-	-	x	-	-	-	-	-	-	-
ta4	Urs-12-en-3β-yl acetate	x	tr	x	x	-	-	-	-	-	-	-
ta5	D:C-friedo-urs-8-en-3β-yl acetate	x	-	-	x	-	-	-	-	-	-	-
ta6	Taraxast-20(30)-en-3β-yl acetate	x	-	-	-	-	-	-	-	-	-	-
ta7	D:A-friedo-olean-3β-yl acetate	x	tr	-	x	-	-	-	-	-	-	-
Pentacyclic Triterpenyl Ketones												
k1	D-friedo-olean-14-en-3-one	x	-	-	x	-	-	-	-	-	-	-
k2	Olean-12-en-3-one	x	x	x	x	-	x	-	-	-	-	-
k3	D:C-friedo-ursan-8-en-3-one	x	tr	x	-	-	-	-	-	-	-	-
k4	D:C-Friedo-5α-olean-7-en-3-one	x	-	-	x	-	-	-	-	-	-	-
k5	Urs-12-en-3-one	x	x	x	x	-	x	-	-	-	-	-
k6	A'-Neogammacer-22(29)-en-3-one	x	-	x	-	-	-	-	-	-	-	-
k7	D:B-friedoolean-5-en-3-one	x	-	-	x	-	-	-	-	-	-	-
k8	Lup-22(29)-en-3-one	x	x	x	x	-	x	-	x	-	-	-
k9	D:B-friedo-B':A'-neogammacer-5-en-3-one	x	-	-	-	-	-	-	-	-	-	-
k10	Fern-7-en-3-one	x	-	-	-	-	x	-	-	-	-	-
k11	D:A-friedo-olean-3-one	x	tr	-	x	-	-	-	-	-	-	-
k12	Lup-20(29)-en-3-one	x	-	x	-	x	-	-	-	-	-	-
Pentacyclic Triterpenols												
tp1	D-friedo-olean-14-ene-3β-ol	x	-	-	x	-	-	-	-	-	-	-
tp2	Olean-12-en-3β-ol	x	x	x	x	-	x	-	-	-	-	-
tp3	Olean-18-en-3β-ol	x	tr	x	x	-	-	-	-	-	-	-
tp4	D:B-friedoolean-5-en-3β-ol	x	-	-	x	-	-	-	-	-	-	-
tp5	Urs-12-en-3β-ol	x	x	x	x	-	x	-	-	-	-	-
tp6	Lup-20(29)-en-3β-ol	x	x	x	x	-	x	-	x	-	-	-
tp7	D:B-friedo-B':A'-neogammacer-5-en-3β-ol	x	-	-	-	-	-	-	-	-	-	-
tp8	D:A-friedo-olean-3β-ol	x	-	-	x	-	-	-	-	-	-	-
Pentacyclic Triterpene Acids												
Oa	3β-Hydroxy-olean-12-en-28-oic acid	x	x	x	x	-	x	-	-	-	-	-
Ua	3β-Hydroxy-urs-12-en-28-oic acid	x	x	x	x	-	x	-	-	-	-	-
Betulin derivatives												
db1	Methyl-3β-acetoxy-lup-22(29)-en-28-oate	x	-	-	-	-	-	-	x	-	-	-
db2	Lupa-2,20(29)-dien-28-ol	x	-	-	-	-	x	-	x	-	-	-

db3	3β-Hydroxy- lup-20(29)-en-28-oic acid	Betulinic acid	x	-	-	-	-	-	x	-	-	-
db4	3β-Lup-20(29)-ene-3,28-diol	Betulin	x	-	-	-	-	-	x	-	-	-

Table 3: Abundances and distribution parameters of *n*-alkyl lipids in soils samples of La Guette peatland.

#	<i>n</i> -Alkanes							<i>n</i> -Alkanols				<i>n</i> -Alkanoic acids			
	<i>Total</i> (μg/g)	<i>major</i> ^a	<i>CPI</i> ^b	<i>ACL</i> ^c	<i>Paq</i> ^d	<i>n</i> -C ₂₃ / <i>n</i> -C ₂₉	<i>(n</i> -C ₃₁ + <i>n</i> -C ₃₃) / (<i>n</i> -C ₂₇ + <i>n</i> -C ₂₉)	<i>Total</i> (μg/g)	<i>major</i> ^a	<i>CPI</i> ^e	<i>ACL</i> ^c	<i>Total</i> (μg/g)	<i>major</i> ^a	<i>CPI</i> ^e	<i>ACL</i> ^c
StI	9.6	<i>n</i> -C ₂₉	2.0	28.8	0.53	0.5	0.7	5.6	<i>n</i> -C ₂₈ OH	23.1	25.6	20.3	<i>n</i> -C _{14:0}	5.7	24.7
StII	16.2	<i>n</i> -C ₃₃	6.5	30.5	0.38	2.3	3.3	11.4	<i>n</i> -C ₂₈ OH	34.3	26.4	91.9	<i>n</i> -C _{26:0}	6.2	25.3
StIII	13.9	<i>n</i> -C ₃₃	2.8	29.6	0.50	0.8	1.4	9.0	<i>n</i> -C ₂₈ OH	37.8	26.3	62.8	<i>n</i> -C _{26:0}	7.8	25.6
StIV	6.0	<i>n</i> -C ₂₅	2.5	27.6	0.61	0.2	0.4	38.0	<i>n</i> -C ₂₈ OH	38.4	26.8	11.9	<i>n</i> -C _{16:0}	7.0	24.5
StV	5.4	<i>n</i> -C ₃₃	3.5	30.2	0.25	0.2	2.1	5.2	<i>n</i> -C ₂₈ OH	39.1	26.7	47.3	<i>n</i> -C _{14:0}	7.2	24.0
StVI	11.1	<i>n</i> -C ₃₃	3.8	30.1	0.28	0.2	1.6	19.8	<i>n</i> -C ₂₈ OH	49.4	27.8	39.3	<i>n</i> -C _{14:0}	7.7	24.2
StVII	8.3	<i>n</i> -C ₃₃	3.5	29.7	0.30	0.3	1.3	12.0	<i>n</i> -C ₂₈ OH	28.2	26.9	12.2	<i>n</i> -C _{22:0}	5.8	24.0
StVIII	5.0	<i>n</i> -C ₂₇	1.7	28.1	0.53	0.4	0.5	7.1	<i>n</i> -C ₂₈ OH	23.5	25.8	24.2	<i>n</i> -C _{14:0}	2.0	24.1
StIX	5.2	<i>n</i> -C ₂₅	1.8	28.5	0.57	0.5	0.7	0.2	<i>n</i> -C ₂₆ OH	4.7	23.4	11.6	<i>n</i> -C _{22:0}	5.0	23.7

Total = total abundance (μg/g of soils).
^aChain length of the most abundant homologue;
^bAlkane carbon preference index; CPI = 2(odd *n*-C₂₃ to *n*-C₃₁)/(even *n*-C₂₂ to *n*-C₃₀+ even *n*-C₂₄ to *n*-C₃₂).
^cAverage chain length, ACL = (Σ[Ci] x i)/Σ[Ci] where i = carbon number ranges (*n*-alkanes from *n*-C₂₃ to *n*-C₃₃; *n*- alkanols and *n*-acids, from C₂₂ to C₃₂), Ci = contents of the homologues containing i carbon atoms.
^dTerrestrial/freshwater plant ratio, Paq = (*n*-C₂₃ + *n*-C₂₅) / (*n*-C₂₇ + *n*-C₂₉ + *n*-C₃₁).
^e*n*-Alkanols and *n*-alkanoic acids, CPI = 2(even C₂₂ to C₃₀)/(odd C₂₁to C₂₉ + odd C₂₃ to C₃₁).